Webcam Geo-localization using Aggregate Light Levels

Nathan Jacobs
University of Kentucky
jacobs@cs.uky.edu

Kylia Miskell, Robert Pless
Washington University in St Louis
{klm4|pless}@cse.wustl.edu

Abstract

We consider the problem of geo-locating static cameras from long-term time-lapse imagery. This problem has received significant attention recently, with most methods making strong assumptions on the geometric structure of the scene. We explore a simple, robust cue that relates overall image intensity to the zenith angle of the sun (which need not be visible). We characterize the accuracy of geo-location based on this cue as a function of different models of the zenith-intensity relationship and the amount of imagery available. We evaluate our algorithm on a dataset of more than 60 million images captured from outdoor webcams located around the globe. We find that using our algorithm with images sampled every 30 minutes, yields localization errors of less than 100km for the majority of cameras.

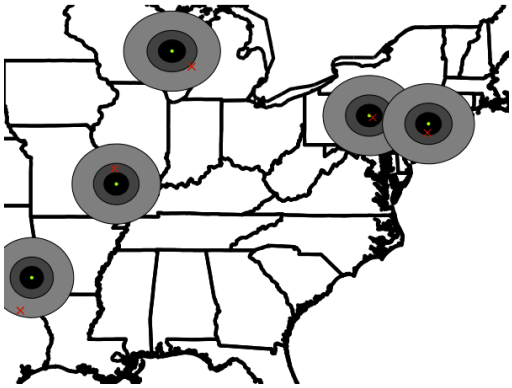


Figure 1: Our algorithm gives location estimates and corresponding, empirically derived, confidence regions for the webcam localization problem. The cross (red) represents the ground-truth location of the camera. The concentric circles represent the 25th, 50th and 75th percentile confidence regions for the estimated camera location (green dot).

1. Introduction

There are currently tens of thousands of public outdoor webcams spread all over the globe, distributing a vast source of imagery every minute. Once cameras are mounted and freely available online they become a general resource for the research community; for example, large sets of traffic, surveillance and resort cameras have been re-purposed for large-scale environmental monitoring [10, 7]. But, without knowing the location of the camera the images are of very little value. This is a problem because most webcams do not provide the camera's location and the accuracy of geo-location estimates based on IP address lookup is often very poor. Given the vast numbers of webcams, automated methods for localizing webcams are an important first step in enabling many uses of the global network of outdoor webcams [10].

Automated algorithms to geo-locate cameras directly from images include approaches based on correlating the mean intensity of the time series with time of day or local cloud cover maps [12], as well as comparing the visible sky to graphics-based models of sky appearance on clear days, first used for camera orientation [11] then for calibration and geo-location [16, 17]. Other systems [13, 3, 21, 5, 19, 8] focus on single-image camera localization and tend to use comparisons between computed image features and test images in a database to pinpoint cameras. Many of these papers present anecdotal results, or only work when the test image sequence is geographically close enough to those images in the training data set. This makes it difficult to extend their results or to be able to predict the accuracy of their geo-location algorithm if it was scaled up to run on the entire global webcam network.

In order to trust and improve automatic geo-location algorithms, more systematic efforts are needed to characterize when they work, when they don't work, and why they don't work. This paper starts this process by considering a formalization of the original algorithms presented in [12], and finding its accuracy as a function of number of days of data and for different forms of the relationship between zenith and intensity.

The focus of this paper is in solving the general webcam

978-1-4244-9497-2/10/$26.00 ©2010 IEEE

geo-location problem [12]: given a series of time-stamped images captured from a static camera, find the geographic location of that camera. We use the known geometry of the sun and earth to explore the relationship of image intensity to solar zenith angle—the angle between a vector normal to the earth's surface and the vector pointing at the sun. We show that this zenith-intensity relationship holds even if the sun is not in the image itself. Using this cue, we provide a general, worldwide solution to the webcam geo-location problem.

1.1. Contributions

Our paper makes several important contributions in the area of automated camera geo-localization. We provide an analysis of the relationship between zenith angle and image intensity for outdoor cameras (Section 3). We also describe a robust algorithm capable of localizing low frame-rate outdoor webcams (Section 4). We provide a maximum likelihood interpretation of the algorithm (Section 4.1). Finally, we evaluate our algorithm on and make publicly available a dataset of webcam imagery; the dataset contains one year of images along with ground-truth locations for over 500 outdoor cameras (Section 5).

2. Related Work

This section provides an overview of the long history of image-based geo-localization algorithms. We categorize algorithms roughly based on the type of cue they use to estimate the location of the camera.

2.1. Localization using Reference Imagery

Reference imagery with a known geographic location is often used for camera localization and is the preferred method when it is available. Finding the position and orientation of cameras in a network where there are feature correspondences between multiple cameras has been extensively studied [13, 3]. Similar reference-based methods, evaluated for the ICCV 2005 Vision Contest, reliably estimated the camera location to within a few meters given a large ground-truth dataset. Methods based on feature correspondences and triangulation are highly accurate but are computationally expensive and require a large ground-truth dataset that is not always available.

In addition to feature correspondences, correlated activity has been explored as a localization cue. Examples include methods based on object tracks [18] and timing correlations of when objects enter and exit the camera views [22]. With these cues, it is possible to estimate the relative camera locations even when the fields of view do not overlap, but the cameras must be nearby and see the same objects.

Larger-scale camera localization using reference imagery is less well studied, but has been approached using feature matching between image features and features computed from a digital elevation map [21, 5, 19]. Recent work [8] has shown that it is possible to compute rough estimates of camera location from a single image without direct feature correspondences by simply finding nearest-neighbors images based on feature similarity.

2.2. Localization using the Sun and Weather

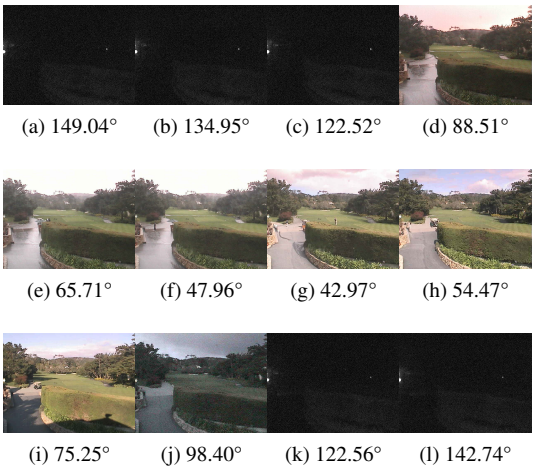
The early work on global camera geo-localization using natural scene variations is based on explicit measurements of the sun position [4]. These techniques require that the sun be visible and that the camera calibration be accurate in order to determine the angle of the sun. Similar constraints have been applied to methods utilizing cast shadows [14, 23]. These shadow-based techniques enable localization when the sun is not visible, but they often make strong assumptions about the scene geometry (e.g. the presence of easy to track shadows moving across a ground plane).

Sunkavalli et. al. [20] used surface color changes induced by the motion of the sun to estimate the camera location and orientation. This work requires intensity measurements with linear camera response. Their experiments with a single camera they found an error of 83.7 km. Our work, while it requires more days worth of imagery, does not expect a linear camera response, works with substantially lower quality imagery, and has comparable accuracy.

Camera localization algorithms based on aggregate light levels has long been used by biologists. For example, the timing of dawn and dusk is often used to estimate the location of tagged marine mammals and fish, such as elephant seals [9]. More recent work has shown improved results by making strong assumptions about the dependence of intensity of a particular spectral band on the time of day [6]. These methods require relatively high frame rates and careful calibration of the imaging hardware in order to ensure an accurate estimate of dawn/dusk times and, therefore, of camera location.

Unfortunately, in the domain of webcam localization, frame rates are often quite low, often only a single frame every 15 minutes, and the precise imaging properties of the camera are unknown. Previous work on webcam localization [12] using light levels finds the geographic location that gives a synthetic intensity sequence (i.e. the zenith sequence modulated by a *fixed* sigmoid function) that is most correlated with the intensity sequence measured by the camera. The use of a fixed modulating function is a significant limitation of this approach; it is problematic because different cameras respond to light differently and the rigid assumption leads to biased location estimates. Our work extends upon this by making weaker assumptions about the modulating function.

We begin the discussion of our algorithm with an em-



(a) 149.04° (b) 134.95° (c) 122.52° (d) 88.51°

(e) 65.71° (f) 47.96° (g) 42.97° (h) 54.47°

(i) 75.25° (j) 98.40° (k) 122.56° (l) 142.74°

Figure 2: A series of images taken two hours apart over the span of one day, with the respective ground-truth zenith angles.

pirical exploration of the relationship between zenith and intensity for cameras with known ground truth.

3. The Relationship between Zenith and Intensity

The zenith angle is the angle formed by a vector pointing straight up, toward the zenith, and a vector pointing at the sun. As such, the zenith angle acts as a measurement of the perceived height of the sun above the horizon at a given location and time. In this section, we explore the relationship between the zenith angle and the scene intensity for outdoor webcams.

Figure 2 shows example images from an outdoor webcam along with the corresponding zenith angle. Clearly, the appearance of the individual pixels due to the zenith angle is determined by a number of factors including: the surface orientation, cast shadows, and scene depth. For camera localization, we find that modeling this complexity is unnecessary: we consider only the average intensity of all the pixels in the scene. The following section describes our empirical observations of the relationship between the time-series of average image intensity and solar zenith angle.

3.1. Observations

Given a set of images, I, each captured at times $T = \{t_1, \ldots, t_n\}$, we define the average intensity of each image as $B = \{b_1, \ldots, b_n\}$. In this section, we describe a collection of empirical observations that help to better understand the distribution, $P(b|z)$, of average intensity values, b, given the zenith angle, z. These observations lead to constraints on the form of the distribution, $P(b|z)$, that enable us to localize an individual camera given a sequence of time-stamped images. To illustrate these observations, we use images from the dataset described in Section 5.1 and derive the zenith angles from the known archive time and camera location.

Figure 3 shows examples of the relationship between zenith and intensity for several real cameras in a broad range of locations. Several properties can be readily observed from such plots. First, there is a strong, consistent, and monotonically decreasing relationship between zenith and intensity: as the sun moves above the horizon (and thus the zenith angle decreases), the intensity of an image increases. Furthermore, the intensity of an image stays roughly constant throughout the day and throughout the night, changing only during dawn and dusk (this is due to automatic exposure control during the day, without this there is a smoother transition). This means that the distribution $P(b|z)$ is similar to a step function with the step approximately corresponding to civil twilight ($zenith = 96\,^\circ$, see Figure 5). This also implies that, for a fixed zenith angle, $\hat{z}$, the distribution $P(b|z = \hat{z})$ has much lower variance than the marginal distribution, except near dawn and dusk. Finally, the variance of the distribution $P(b|z)$ is dependent on the correctness of geo-location estimate. That is, an incorrect geo-location will give distributions $P(b|z)$ with much higher conditional variance, as shown in Figure 3.

In this section, we described common properties of the distribution of average intensity given zenith angle, $P(b|z)$, for different camera and different geographic locations. Next, we show how to leverage these properties of the zenith-intensity relationship to estimate the geographic location of an outdoor camera.

4. Webcam Localization using Light Levels

We have shown that there is a strong and predictable relationship between zenith and intensity for outdoor webcams. We use this to motivate an algorithm for estimating the location of a static outdoor camera. Essentially, the algorithm searches for the geographic location that gives the distribution of average intensities, $P(b|z)$, that is most consistent with the observations in the previous section. The remainder of this section describes and formalizes our approach.

4.1. MLE Formulation of Light-Based Geo-localization

Our goal is to determine the geographic location, L^*, of the camera given only the intensity sequence, B, and the corresponding time stamps, T. We model this in the maximum-likelihood estimation framework. This yields the following initial optimization problem:

$$L^* = \arg\max_L P(B|\Theta, L, T), \tag{1}$$

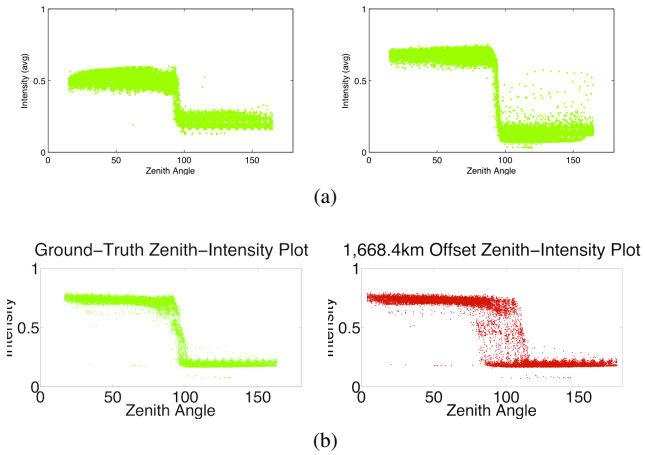


Figure 3: (a) Examples of the relationship between zenith and intensity for two cameras. (b) An example of the relationship for the correct (left) and the incorrect (right) camera location. In this case the ground truth was adjusted by 1,500 km prior to computing the zenith angle.

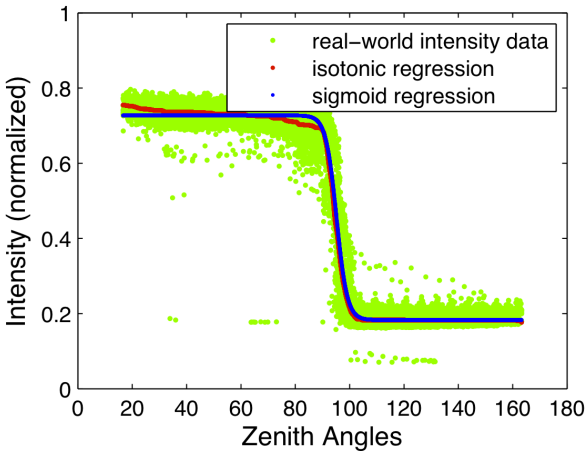


Figure 4: The scatter plot of intensity values for a single camera along with the corresponding regression models. The red curve corresponds to the isotonic regression model and the blue curve to the sigmoid. Note that while the monotonic regression model gives a a better estimate of the expected value (especially at left) the sigmoid model is quite accurate.

where Θ represents the parameters of the camera and scene properties that affect the intensity. These properties include the radiometric response function, the automatic exposure compensation method, and the buildings and trees that shadow the area viewed by the camera. In this work we ignore this complexity and make the assumption that the average intensity decreases as the sun is lower in the sky. The parameters, Θ, specify the expected value of the distribution and depend on the specific regression model being used. In this work, we use either a piece-wise linear or a sigmoid model (see Section 4.2 for details).

We note that Equation 1 can be simplified because zenith is a well-known function of location and time, then write the problem as:

$$L^* = \arg\max_L P(B|\Theta, z(L,T)) \tag{2}$$

We assume conditional independence amongst the samples in B and rewrite the problem as:

$$L^* = \arg\max_L \prod_{i=1}^{n} P(b_i|\Theta, z(L,t_i)). \tag{3}$$

Since we are not interested in the values Θ, we replace this variable with the optimal value, $\Theta^*(L)$, at a given location:

$$\Theta^*(L) = \arg\max_\Theta \prod_{i=1}^{n} P(b_i|\Theta, z(L,t_i)). \tag{4}$$

Our methods for parametrizing Θ are described in Section 4.2. This yields the following problem:

$$L^* = \arg\max_L \prod_{i=1}^{n} P(b_i|\Theta^*, z(L,t_i)) \tag{5}$$

Our method for optimizing this problem are given in Section 4.3.

Several aspects of the distribution, $P(b_i|\Theta, z(L,t_i))$, have been left unspecified: the error model, and the parametrization of Θ. For the error model, we assume that intensity is conditionally Laplacian. We found that this gave superior results to a Gaussian assumption, likely due to better handling of outlier intensity values. In our dataset, outliers are caused by many factors, including lens flares, myriad sensor abnormalities, and specularities.

The next section describes the final important aspect of our model: the dependence on Θ.

4.2. Parametrizing the Relationship between Zenith and Intensity

In our model, Θ represents the unknown camera and scene parameters that determine the form of the distribution, $P(b_i|\Theta, z(L,t_i))$. More specifically, the parameters, Θ, determine the expected value of intensity given zenith angle. In this section, we describe two parametrizations of Θ: a low-dimensional sigmoid model and a non-parametric, monotonic regression model.

The sigmoid model has four parameters, $(\theta_1, \ldots, \theta_4)$, which determine the expected value of intensity given the

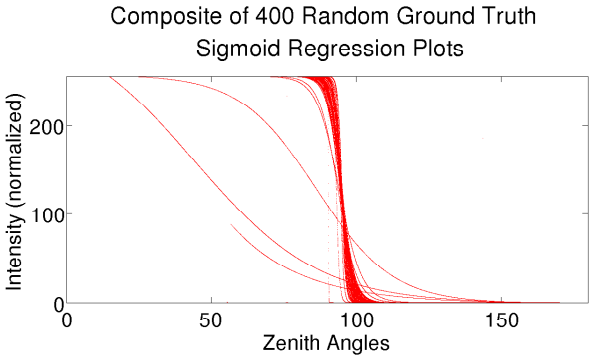


Figure 5: A sigmoidal model of the zenith-intensity relationship fit separately to data from 400 different cameras (see Section 4.2). This highlights that the strong transition from bright to dark often occurs at $zenith = 96\,^{\circ}$.

zenith angle. The model is as follows:

$$E(b|\Theta, z) = \theta_1 - \frac{\theta_2}{1 + e^{\theta_3(\theta_4 - z)}}. \quad (6)$$

This model has a small number of parameters, and is easy to optimize, but is not as flexible as the non-parametric model. Figure 4 shows an example of the sigmoid regression model fit to a collection of sample intensity values. Extending this to many more cameras, Figure 5 shows examples of this model fit to data from 400 different cameras. This highlights that the transition from dawn to dusk occurs at roughly the same zenith angle for all cameras.

The monotonic regression model [15] uses linear programming to solve for a piecewise-linear, monotonically decreasing function of zenith given intensity. We regularly sample the zenith angle over the interval $[0, 180]$ and solve for the values of intensity at these points that simultaneously minimize the L_1-norm residual and are monotonically decreasing. We find that 100 sample points is sufficiently flexible and not too computationally demanding. Figure 4 shows an example of the monotonic regression model fit to a collection of sample intensity values. This highlights aspects of the expected value that cannot be model with the low-dimensional sigmoid model.

We then reformulate the objective function, in the standard fashion, to minimize the negative log-likelihood. This gives the following, and final, optimization problem:

$$L^* = \arg\min_{L} \sum_{i=1}^{n} \|b_i - E(b|\Theta^*, z(L, t_i))\|_1 \quad (7)$$

In the following section, we show how we optimize Equation 7.

4.3. Finding the Optimal Location

Given the low-dimensionality of the optimization problem, we solve for the optimal location, L^*, using the following simple method. First, we sample a large number of locations (typically a grid of several hundred points). We then select the top-k solutions from the grid and locally optimize using Nelder-Mead simplex search (we find that $k = 10$ initial locations is sufficient). The lowest-error solution after the simplex search is our estimate of the camera location.

This simple procedure uses robust properties of the relationship between zenith and intensity. We find that our unoptimized implementation requires approximately a minute to estimate the location of a camera, given 10 000 images, on a standard desktop computer (ignoring the time required to compute the image averages). In the following evaluation section, we show the performance of this algorithm via an extensive set of experiments.

5. Evaluation

We evaluated our method on an extensive real-world dataset, this section describes the dataset and the interesting aspects of this evaluation. We begin with a description of the dataset.

5.1. An Evaluation Dataset

The data we use for evaluation [2] was captured from real outdoor webcams located around the globe; it is a subset of the Archive of Many Outdoor Scenes (AMOS) dataset [1]. The AMOS dataset consists of 60 million images captured from webcams since March 2006. In addition to the image data, the dataset provides the timestamps at which the images were archived and ground-truth geographic location for some of the cameras. The ground-truth locations were determined by a variety of methods, including manually searching Google Maps for corresponding points. Locations of the cameras we used are shown in Figure 6. To the best of our knowledge, this is the largest public data set that provides time-stamped imagery from many cameras distributed across the globe, their respective time-stamps, and known geo-locations.

5.2. Localization Error by Regression Model

In this section, we show the results of an extensive evaluation on 365 days of images from 633 unique outdoor cameras. We compare the two regression methods and find that they work similarly, but with the sigmoid model clearly dominating the non-parametric model. Figure 7 summarizes the results as cumulative distribution functions of errors. We find that 75% of the cameras are geo-located within about 200 km of the ground-truth location, no matter what regression model is used. Furthermore, about 50% of

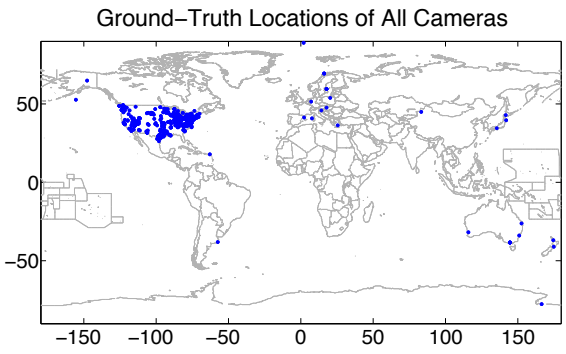


Figure 6: The locations of the 633 cameras we use to evaluate our algorithm.

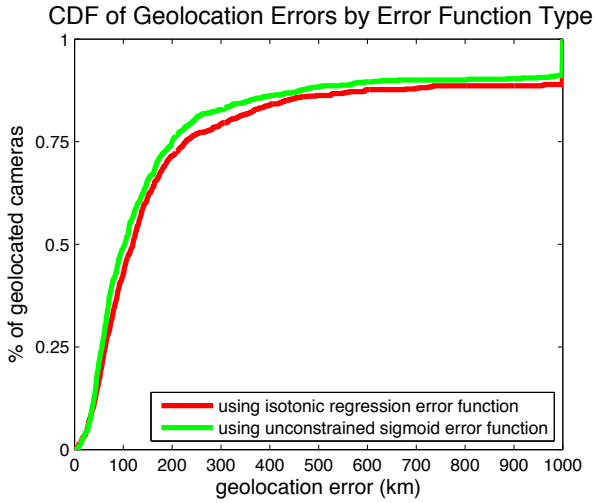


Figure 7: The cumulative distribution of errors of two regression models. Errors are estimated from a dataset of 633 cameras with known ground-truth locations.

all cameras are geo-located within 100 km, no matter what regression model is used. However, the sigmoid regression model performs better on the cameras with high localization error: while the using the sigmoid regression model geo-locates 90% of all cameras within approximately 350 km of ground truth, the monotonic regression model has an error of 580 km.

Given the similarity in performance, we use the simpler sigmoid model in the following section.

5.3. Localization Error by Number of Days

In addition to depending on the regression model, the localization error depends on the number of available days of imagery. In this experiment we use the sigmoid regression model and solve for the camera location for temporal ranges from 5 to 365 days. Figure 8 shows the 25th, 50th, and 75th percentile error for each number of days. From this experiment we observe that using more than 90 days of images does not significantly improve localization accuracy.

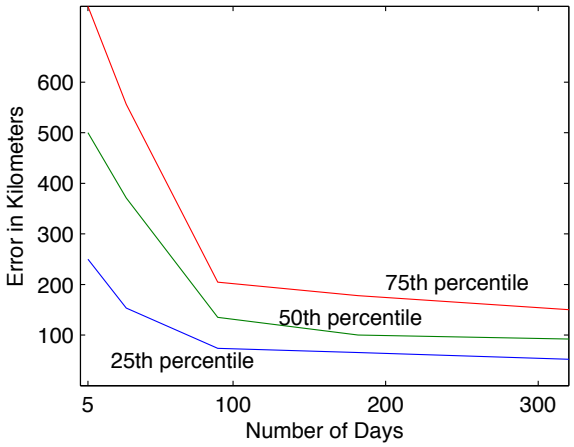


Figure 8: The localization accuracy depends on the number of days of imagery used. This plot shows the dependence of the 25th, 50th, and 75th percentile error on the number of days for a dataset of 50 cameras. From this we see that this method reaches a point of diminishing returns at approximately 90 days.

6. Discussion

Other work has shown promising results using a variety precise geometrical localization cues. The challenge with these methods is that the often depend on high-quality image data and specific scene configurations. This limits the usefulness of these methods, and may, in part, explain why these methods are often evaluated on a small number of scenes. Instead, this work explores a much simpler cue, the relationship between average image intensity and the zenith angle of the sun.

Our use of the zenith angle as a cue for geo-localization relies on the empirical relationship between the intensity of a outdoor scene and the sun's height above the horizon. This requires us to make certain assumptions: we assume that this relationship is invariant to the potentially time-varying camera parameters and weather conditions, we assume the intensity of the scene monotonically decreases as the zenith angle increases (ignoring the possibility of heavy shadow during the day or the presence of man-made lighting during the night), and we assume that camera images are correctly time-stamped (incorrect time-stamps lead to incorrect longitude estimates). Despite these limitations, our experiments on a real-world, unfiltered dataset, show that the majority of webcams can be localized to within 100km using our method.

7. Conclusion

Solving the webcam geo-localization problem is an important first step in using the global network of outdoor webcams. Existing methods for camera geo-localization do not work well in this domain for a variety of reasons, largely due to strong assumptions about the contents of the scene. We presented an approach to solving this problem that relies on a simple, easy-to-measure cue that relate the solar zenith angle to the image intensity. In addition, we evaluated on a large dataset of webcam imagery which we are making publicly available.

References

[1] AMOS website. http://amos.cse.wustl.edu/.

[2] Webcam geo-localization dataset. http://cs.uky.edu/~jacobs/research/geolocate/.

[3] P. Baker and Y. Aloimonos. Calibration of a multicamera network. In *Omnivis 2003: Omnidirectional Vision and Camera Networks*, 2003.

[4] F. Cozman and E. Krotkov. Robot localization using a computer vision sextant. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, pages 106–111, Nagoya, Japan, May 1995.

[5] F. Cozman and E. Krotkov. Automatic mountain detection and pose estimation for teleoperation of lunar rovers. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, 1997.

[6] P. Ekstrom. An advance in geolocation by light. *Memoirs of the National Institue of Polar Research, Special Issue. National Institute of Polar Research, Tokyo*, pages 210–226, 2004.

[7] E. A. Graham, E. C. Riordan, E. M. Yuen, J. Hicks, E. Wang, D. Estrin, and P. W. Rundel. Leveraging internet-connected cameras to create a transcontinental plant phenology monitoring system. Ecological Society of America (ESA) Annual Meeting, Aug. 2009.

[8] J. Hays and A. A. Efros. im2gps: estimating geographic information from a single image. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[9] R. D. Hill. Theory of geolocation by light levels. In B. J. L. Boeuf and R. M. Laws, editors, *Elephant seals: population ecology, behavior and physiology*, pages 227–236. University of California Press, 1994.

[10] N. Jacobs, W. Burgin, N. Fridrich, A. Abrams, K. Miskell, B. H. Braswell, A. D. Richardson, and R. Pless. The global network of outdoor webcams: Properties and applications. In *ACM SIGSPATIAL*, Nov. 2009.

[11] N. Jacobs, N. Roman, and R. Pless. Toward fully automatic geo-location and geo-orientation of static outdoor cameras. In *Proceedings IEEE Workshop on Applications of Computer Vision*, 2008.

[12] N. Jacobs, S. Satkin, N. Roman, R. Speyer, and R. Pless. Geolocating static cameras. In *Proceedings IEEE International Conference on Computer Vision*, 2007.

[13] J. Jannotti and J. Mao. Distributed calibration of smart cameras. In *Workshop on Distributed Smart Cameras*, 2006.

[14] I. Junejo and H. Foroosh. Estimating geo-temporal location of stationary cameras using shadow trajectories. In *Proceedings European Conference on Computer Vision*. 2008.

[15] J. Kruskal. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 29(2), June 1964.

[16] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What does the sky tell us about the camera? In *Proceedings European Conference on Computer Vision*, 2008.

[17] J.-F. Lalonde, S. G. Narasimhan, and A. A. Efros. What do the sun and the sky tell us about the camera? *Int. J. Comput. Vision*, 88(1):24–51, 2010.

[18] D. Marinakis and G. Dudek. Topology inference for a vision-based sensor network. In *Canadian Conference on Computer and Robot Vision (CRV)*, pages 121–128, 2005.

[19] F. Stein and G. Medioni. Map-based localization using the panoramic horizon. In *Proc. IEEE International Conference on Robotics and Automation (ICRA)*, Nice, France, 1992.

[20] K. Sunkavalli, F. Romeiro, W. Matusik, T. Zickler, and H. Pfister. What do color changes reveal about an outdoor scene? In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[21] W. Thompson, T. Henderson, T. Colvin, L. Dick, and C. Valiquette. Vision-based localization. In *ARPA Image Understanding Workshop*, pages 491–498, Washington D.C., 1993.

[22] K. Tieu, G. Dalley, and W. E. L. Grimson. Inference of non-overlapping camera network topology by measuring statistical dependence. In *Proceedings IEEE International Conference on Computer Vision*, pages 1842–1849, 2005.

[23] L. Wu, X. Cao, and H. Foroosh. Camera calibration and geo-location estimation from two shadow trajectories. 114(8):915–927, 2010.